

A reduced-order derivative-informed neural operator for subsurface fluid-flow

Jeongjin Park, Grant Bruer, Huseyin Tuna Erdinc, Abhinav Prakash Gahlot, and Felix J. Herrmann
Georgia Institute of Technology

SUMMARY

Neural operators have emerged as cost-effective surrogates for expensive fluid-flow simulators, particularly in computationally intensive tasks such as permeability inversion from time-lapse seismic data, and uncertainty quantification. In these applications, the fidelity of the surrogate’s gradients with respect to system parameters is crucial, as the accuracy of downstream tasks, such as optimization and Bayesian inference, relies directly on the quality of the derivative information. Recent advances in physics-informed methods have leveraged derivative information to improve surrogate accuracy. However, incorporating explicit Jacobians can become computationally prohibitive, as the complexity typically scales quadratically with the number of input parameters. To address this limitation, we propose DeFINO (Derivative-based Fisher-score Informed Neural Operator), a reduced-order, derivative-informed training framework. DeFINO integrates Fourier neural operators (FNOs) with a novel derivative-based training strategy guided by the Fisher Information Matrix (FIM). By projecting Jacobians onto dominant eigen-directions identified by the FIM, DeFINO captures critical sensitivity information directly informed by observational data, significantly reducing computational expense. We validate DeFINO through synthetic experiments in the context of subsurface multi-phase fluid-flow, demonstrating improvements in gradient accuracy while maintaining robust forward predictions of underlying fluid dynamics. These results highlight DeFINO’s potential to offer practical, scalable solutions for inversion problems in complex real-world scenarios, all at substantially reduced computational cost.

INTRODUCTION

Employing neural networks as surrogate models to approximate solutions of parametric PDEs has shown promise in addressing many-query scenarios such as Bayesian inference, optimization under uncertainty, and Bayesian optimal experimental design (Qiu et al., 2024; O’Leary-Roseberry et al., 2024). Such problems typically rely on complex mathematical models, often parameterized PDEs, which require numerous forward model evaluations, i.e., solving the PDEs for many parameter settings, making high-fidelity simulation computationally demanding (Gao et al., 2024; Sherifdeen et al., 2019; Wu et al., 2024). To overcome this challenge, neural operator surrogates, particularly Fourier Neural Operators (FNOs), have emerged as efficient alternatives for approximating PDE solutions (Yin et al., 2022; Li et al., 2020a; Wen et al., 2022; Chandra et al., 2025). By leveraging computationally inexpensive approximations of the forward model, these surrogates significantly reduce the computational burden, thus enabling more scalable and practical downstream tasks, such as inversion problems, and uncertainty quantification (Yin et al., 2023).

When neural operators are used as surrogates for numerical simulators or governing equations of physical systems, it is essential that they adhere to fundamental physical principles, such as mass conservation, entropy decay, and invariant measures (Li et al., 2024; Park et al., 2024; Cai et al., 2021). However, standard neural network training approaches, typically based on mean squared error (MSE) objectives, may struggle to accurately learn true physics, limiting the reliability and generalizability of the resulting surrogate models (Park et al., 2024; Rosofsky et al., 2023; Li et al., 2024). For instance, FNOs have exhibited instabilities in long-term predictions, particularly for stiff systems, systems that evolve on vastly different time scales, highlighting the need for robust training strategies capable of faithfully capturing complex physical behaviors (Park et al., 2024; Li et al., 2024).

Such inaccuracies can significantly compromise the robustness and reliability of subsequent downstream tasks, such as Bayesian inversion, optimization under uncertainty, and Bayesian optimal experimental design, for which a precise gradient is critical. To address these limitations, particularly for inherently complex problems such as chaotic systems (e.g., weather forecasting) or high-dimensional time-dependent fluid-flow scenarios, recent studies have introduced derivative-based training approaches. Approaches such as Physics-Informed Neural Networks (PINNs), and Derivative-Informed Neural Operators (DINOs) integrate derivative-based penalty terms, PDE residual constraints, and higher-order derivative information directly into the training process (Li et al., 2024; Park et al., 2024; O’Leary-Roseberry et al., 2024; Wu et al., 2023). The primary motivation behind these derivative-informed strategies is to enhance surrogate models’ ability to accurately represent not only input-output relationships but also underlying sensitivities and dynamical behaviors governed by PDE systems (O’Leary-Roseberry et al., 2022). Nevertheless, applying these methods to complex, high dimensional real-world problems, such as subsurface fluid-flow, remains challenging due to stringent computational requirements, including fully differentiable simulators and substantial memory and runtime demands (O’Leary-Roseberry et al., 2024).

Learning derivatives with reduced-order modeling

One of the major obstacles in physics-based machine learning, particularly for high-dimensional systems, is the high computational cost required to compute derivatives. This renders many conventional derivative estimation methods impractical for large-scale, real-world problems (O’Leary-Roseberry et al., 2022). Thus, researchers have employed dimension reduction techniques, such as Principal Component Analysis (PCA) or Active Subspace, to compress the input or output into an efficient lower-dimensional representation, resulting in so-called *reduced-order modeling* (O’Leary-Roseberry et al., 2022; Constatine et al., 2014; Benner et al., 2015; Quarteroni et al., 2015). However, despite their advantages, reduced-order mod-

els, which reduce the dimensions of input-output, may encounter limitations when applied to highly nonlinear or multi-scale systems (Constantine, 2015; Benner et al., 2015; Quarteroni et al., 2015). That is, such approaches risk discarding critical information if essential features are not captured within the reduced subspace, which may lead to inaccurate derivative approximations.

Our approach addresses these limitations by computing matrix-free actions of the Jacobian without reducing the dimensionality of the input or output, thus maintaining the full-order structure of the problem. That is, our training algorithm encourages the neural operator to approximate full-order Jacobians in a scalable manner by leveraging eigenvectors derived from the Fisher Information Matrix (FIM), as defined in Definition 1.

Definition 1 (Fisher Information Matrix). *Let \mathbf{A} denote the input parameters and \mathbf{Y} represent the observations. Given a likelihood function $p(\mathbf{Y} | \mathbf{A})$, the Fisher Information Matrix (FIM), denoted by $I(\mathbf{A})$, is defined as:*

$$I(\mathbf{A}) = \mathbb{E}_{\mathbf{Y}} \left[(\nabla_{\mathbf{A}} \log p(\mathbf{Y} | \mathbf{A}))^\top (\nabla_{\mathbf{A}} \log p(\mathbf{Y} | \mathbf{A})) \right]. \quad (1)$$

Unlike traditional data-driven dimension reduction methods such as Principal Component Analysis (PCA) or Proper Orthogonal Decomposition (POD), the FIM offers a gradient-based approach directly informed by the likelihood function (Quarteroni et al., 2015). Furthermore, in contrast to other gradient-based methods like Active Subspace, which typically rely on gradient evaluations obtained by sampling the parameter space according to a predefined prior, the FIM inherently incorporates observational uncertainty through the likelihood function itself (Constantine et al., 2014), (Zahm et al., 2022), (Cui and Zahm, 2021). Consequently, in our context for the subsurface fluid-flow problem, where the underlying distribution of permeability models is unknown (Gahlot et al., 2024), the FIM provides a more suitable and robust basis for dimension reduction.

Our contribution

To address these challenges, we propose a Derivative-based Fisher-score Informed Neural Operator (DeFINO), a novel framework designed to enhance neural operators by explicitly incorporating reduced-order Jacobian information into the training process. DeFINO differs from recent derivative-informed modeling approaches (O’Leary-Roseberry et al., 2022; O’Leary-Roseberry et al., 2024; Zahm et al., 2022; Qiu et al., 2024) by maintaining the surrogate model in its full-order representation, rather than reducing the input or output dimensionality directly. Instead, DeFINO achieves dimension reduction by projecting the Jacobian onto dominant eigenvectors of the FIM. This strategy focuses the surrogate model’s training on the parameter directions most sensitive to observational data, enabling accurate yet computationally efficient derivative approximations, particularly beneficial for high dimensional problems.

Our primary contributions include: (1) introducing a novel derivative-informed training strategy for neural operators applicable to high-dimensional systems; (2) developing a scal-

able method to compute the FIM, including deriving a closed-form expression for the practical scenario of a multivariate Gaussian likelihood; and (3) validating the effectiveness of the DeFINO framework through numerical experiments. We specifically evaluate DeFINO’s performance on a subsurface fluid-flow problem, highlighting its effectiveness even when the available number of training samples is limited.

METHODOLOGY

Problem description: empirical risk minimization

In this work, our goal is to improve the surrogate model’s prediction of subsurface fluid-flow behavior while accurately capturing gradients with respect to the input permeability field. The plume dynamics are modeled by two-phase flow equations, which we express as the mapping $\mathbf{A} \rightarrow \mathbf{U} = \mathcal{F}(\mathbf{A})$, where \mathcal{F} represents the two-phase flow simulation that models the time-lapse of CO₂ concentration snapshots $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k]$ given the permeability field \mathbf{A} . Our proposed training strategy is formalized the following empirical risk minimization problem:

$$\min_{\theta} \frac{1}{N} \sum_{j=1}^N \left[\left\| \mathcal{F}(\mathbf{A}_j) - \mathcal{F}_{nn}(\mathbf{A}_j; \theta) \right\|_2^2 + \lambda \sum_{i=1}^r \left\| \mathbf{v}_i^\top \mathbf{J}_{\mathcal{F}_{nn}}(\mathbf{A}_j; \theta) - \mathbf{v}_i^\top \mathbf{J}_{\mathcal{F}}(\mathbf{A}_j) \right\|_2^2 \right] \quad (2)$$

where $\mathbf{J}_{\mathcal{F}}(\mathbf{A}) = \frac{\partial \mathcal{F}(\mathbf{A})}{\partial \mathbf{A}}$ and $\mathbf{J}_{\mathcal{F}_{nn}}(\mathbf{A}) = \frac{\partial \mathcal{F}_{nn}(\mathbf{A})}{\partial \mathbf{A}}$ represent the Jacobians of the reservoir simulator and its neural network surrogate, respectively. The vectors $\{\mathbf{v}_i\}_{i=1}^r$ serve as the probing directions derived from the FIM, with r specifying the number of rank. Finally, λ is our regularization parameter. This objective function combines the ℓ_2 -norm misfit of the forward solutions between the reservoir simulator and the surrogate FNO, along with the ℓ_2 -norm misfit of the corresponding vector-Jacobian product. By jointly minimizing these terms over N training samples, we drive the neural operator toward improved accuracy in both forward predictions and gradient estimation.

Fourier Neural Operator (Li et al., 2020a)

Among the various neural operators, such as the Graph Neural Operator (GNO) and Low-Rank Neural Operator (LNO), the Fourier Neural Operator (FNO) stands out for its computational efficiency (Li et al., 2020b,c). By learning weights in the Fourier space, FNO transforms convolutional operations into simple multiplications, thereby significantly accelerating the solution of PDEs. This efficiency makes FNO a popular surrogate model for data-driven approaches to solving numerical PDEs (Lu et al., 2021; Kovachki et al., 2023; Karniadakis et al., 2021). Especially, despite the initial training cost, FNO has been shown to generate approximate PDE solutions orders of magnitude faster than traditional numerical solvers (Li et al., 2020a). That is, given an unseen spatial distribution of permeability \mathbf{A} , the trained FNO can near instantly produce time-dependent plume concentration snapshots via forward evaluation, i.e., $\mathcal{F}_{nn}(\mathbf{A})$ (Li et al., 2020b; Wen et al., 2022; Zhang

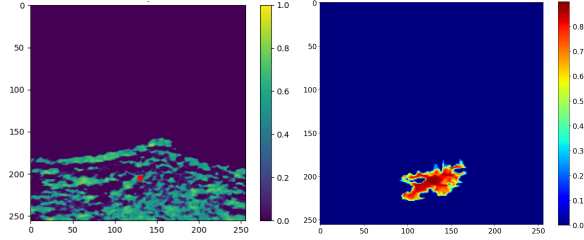


Figure 1: (Left) Example of a heterogeneous permeability model, \mathbf{A} , with injection location shown as red dot. Adapted from (Gahlot et al., 2024) (Right) CO_2 plume concentration after 1,825 days of injection.

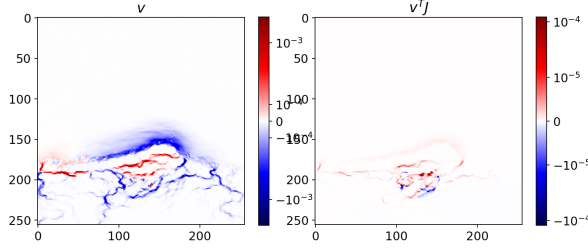


Figure 2: (Left) First eigenvector \mathbf{v} of the Fisher Information Matrix (FIM) for a sample permeability model after 1,095 days of CO_2 plume evolution. (Right) Corresponding ground-truth vector-Jacobian product, $\mathbf{v}^\top \mathbf{J}$, computed using the reservoir simulator \mathcal{F} .

et al., 2022). Moreover, automatic differentiation provides access to gradients with respect to the FNO’s input, further supporting downstream optimization tasks.

Creating joint samples

In the offline phase, we generate joint samples $\{\mathbf{A}_j, \mathbf{U}_j, \mathbf{v}_j, \mathbf{v}_j^\top \mathbf{J}\}_{j=1}^N$ to train our surrogate model. Here, $\mathbf{U} = \mathcal{F}(\mathbf{A})$ with $\mathbf{A}, \mathbf{U} \in \mathbb{R}^{d \times d}$, where d denotes the number of grid points along each spatial dimension. The mapping \mathcal{F} denotes the fluid-flow model implemented in JutulDarcy.jl (Møyner, 2024) and JutulDarcyRules.jl (Yin et al., 2024). For each sample, \mathbf{v}_i is the i th eigenvector of the FIM, and $\mathbf{v}_i^\top \mathbf{J}$ is the corresponding vector-Jacobian product.

A major challenge in this process is the high computational cost of forming the FIM, which can sometimes be intractable. To address this, we consider an observation model

$$\mathbf{Y} = \mathcal{F}(\mathbf{A}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

which yields the following closed-form expression for the Fisher Information Matrix (FIM):

$$\mathbf{I}(\mathbf{A}) = \left(\frac{\partial \mathcal{F}(\mathbf{A})}{\partial \mathbf{A}} \right)^\top \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \mathcal{F}(\mathbf{A})}{\partial \mathbf{A}} \right), \quad (3)$$

where $\mathcal{F} : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$ is a deterministic mapping from the input permeability field to the CO_2 concentration.

To compute the first r eigenvectors of $\mathbf{I}(\mathbf{A})$, we extract the left singular vectors of \mathbf{J} , noting that the eigenvectors of $(\mathbf{v}^\top \mathbf{J})(\mathbf{v}^\top \mathbf{J})^\top$

coincide with the left singular vectors of $\mathbf{v}^\top \mathbf{J}$. This relies on the assumption that, in high-dimensional systems, the gradient information lies intrinsically in a low-dimensional subspace—making the Gauss–Newton Hessian $\mathbf{J}^\top \mathbf{J}$ amenable to low-rank approximation (O’Leary-Roseberry et al., 2024).

Examples of \mathbf{v} and the corresponding vector-Jacobian product $\mathbf{v}^\top \mathbf{J}$, computed using the proposed approach, are illustrated in Figure 2. The plot shows the first eigenvector \mathbf{v} of the FIM and the corresponding Jacobian action $\mathbf{v}^\top \mathbf{J}$, based on the permeability model and CO_2 plume shown in Figure 1. While \mathbf{v} reveals the sensitivity structure with respect to the permeability field, $\mathbf{v}^\top \mathbf{J}$ highlights regions where the saturation has evolved, concentrating around the plume front. This approach avoids explicitly forming the full Gauss–Newton Hessian $\mathbf{J}^\top \mathbf{J}$, which would require $O(d^4)$ operations, and instead computes only r vector-Jacobian products, reducing both time and memory complexity to $O(r \times d^2)$.

NUMERICAL EXPERIMENTS

In our numerical experiments, we validate the performance of our proposed framework within a multiphase flow scenario by predicting the evolution of plume concentration from a given permeability field, represented through time-lapse snapshots. We assess both the output prediction accuracy and the fidelity of the directional derivative, $\mathbf{v}^\top \mathbf{J}$, as computed by the neural operator.

Training setup

In this work, we focus on evaluating the efficacy of our framework with a limited training dataset, consisting of 64 pairs of heterogeneous permeability fields (shown in Figure 1) and the corresponding time evolution of CO_2 concentration (example snapshot shown in Figure 1). The permeability models are defined on a 256×256 grid with a grid spacing of 15 m in both the vertical and horizontal directions, and the associated porosity models are both derived from the work of (Gahlot et al., 2024). For the fluid-flow simulations, an injection well is placed on the middle part of the model, injecting supercritical CO_2 into the reservoir with the injection rate of $7 \times 10^{-3} \text{ m}^3/\text{s}$. The time evolution of CO_2 concentration is modeled using JutulDarcy.jl (Møyner, 2024) and JutulDarcyRules.jl (Yin et al., 2024) over a period of 5 years, with a time step of 1 year. This simulation produces five distinct CO_2 plume snapshots at different time steps for each permeability model. The Fourier Neural Operator (FNO) implementation employed in this study is based on the NVIDIA Modulus framework.

Results and evaluation metrics

After training both models under the same hyperparameter setting, we evaluate the effectiveness of the proposed DeFINO framework by assessing both forward prediction accuracy and gradient accuracy on a set of unseen test samples. Each test sample consists of a permeability model and five corresponding snapshots of CO_2 concentration obtained at intervals of 365 days using reservoir simulator. Following the procedure outlined in the Methodology section, form a test set comprising $\{\mathbf{A}_k, \mathbf{U}_k, \mathbf{v}_k, \mathbf{v}_k^\top \mathbf{J}\}_{k=1}^{100}$.

	FNO	DeFINO
ℓ_2 -norm Misfit of Forward	2.1440×10^{-3}	2.0725×10^{-3}
Relative ℓ_2 -norm Misfit of Gradient	1.6426×10^{-5}	1.1739×10^{-6}
Reduced Gauss-Newton Error	0.8554	0.7697

Table 1: Comparison of FNO and DeFINO in gradient accuracy. Relative ℓ_2 -norm Misfit of Gradient is defined in Equation 4. Reduced Gauss–Newton Error (Equation 5) measures the relative discrepancy (in a reduced subspace spanned by \mathbf{V}_r) between the true and predicted Gauss–Newton Hessians (Adapted from (O’Leary-Roseberry et al., 2024)). Bold values indicate lower (i.e. better) errors.

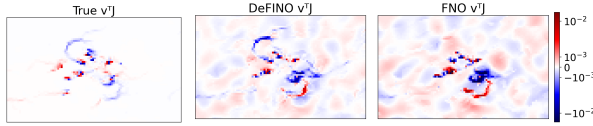


Figure 3: Close-up view of the gradient $\mathbf{v}^\top \mathbf{J}$ after 365 days of CO_2 injection. The first column shows the ground-truth gradient, the second column shows the gradient predicted by DeFINO, and the third column shows the gradient predicted by the baseline FNO. The relative ℓ_2 -norm misfit of DeFINO’s $\mathbf{v}^\top \mathbf{J}$ shown in the middle column is 0.0446, compared to 0.1470 for FNO shown in the right column.

We first compare the forward inference performance of a baseline FNO, trained solely with mean squared error (MSE) loss, against our proposed DeFINO model. For CO_2 plume prediction, DeFINO achieves a slightly improved ℓ_2 -norm misfit of 2.0725×10^{-3} compared to the baseline FNO’s misfit of 2.1440×10^{-3} . This improvement in forward prediction suggests that incorporating gradient information during training enhances not only the model’s ability to learn gradients but also its accuracy in capturing fluid plume evolution.

Beyond forward predictions, we also evaluate gradient accuracy, which is critical in inversion problems and uncertainty quantification tasks (Yin et al., 2023; O’Leary-Roseberry et al., 2022). Specifically, we assess the accuracy of the surrogate-derived gradient, $\mathbf{v}^\top \mathbf{J}_{\mathcal{F}_{nn}}$, compared to the true gradient computed via the reservoir simulator, $\mathbf{v}^\top \mathbf{J}_{\mathcal{F}}$, across the test samples. We quantify this accuracy using two metrics: the relative ℓ_2 -norm misfit of the gradients (Equation 4) and the reduced Gauss–Newton error (Equation 5), which measures how well the surrogate approximates the true Gauss–Newton Hessian ($\mathbf{J}^\top \mathbf{J}$) along dominant parameter directions identified by the FIM (O’Leary-Roseberry et al., 2024). Gauss–Newton Hessians are particularly valuable in inversion and optimization settings, as they provide curvature information essential for robust parameter estimation (O’Leary-Roseberry et al., 2022).

$$\frac{\|\mathbf{V}^\top \mathbf{J}_{\mathcal{F}} - \mathbf{V}^\top \mathbf{J}_{\mathcal{F}_{nn}}\|_F^2}{\|\mathbf{V}^\top \mathbf{J}_{\mathcal{F}}\|_F^2} \quad (4)$$

$$\sqrt{\mathbb{E}_{\mathbf{A}} \left[\frac{\|\mathbf{V}_r^\top (\mathbf{J}_{\mathcal{F}}^\top \mathbf{J}_{\mathcal{F}} - \mathbf{J}_{\mathcal{F}_{nn}}^\top \mathbf{J}_{\mathcal{F}_{nn}}) \mathbf{V}_r\|_F^2}{\|\mathbf{V}_r^\top \mathbf{J}_{\mathcal{F}}^\top \mathbf{J}_{\mathcal{F}} \mathbf{V}_r\|_F^2} \right]} \quad (5)$$

Our results, summarized in Table 1, demonstrate that DeFINO significantly improves gradient estimation accuracy, achieving a relative ℓ_2 -norm misfit (Equation 4) of 1.1739×10^{-6} compared to the baseline FNO’s misfit of 1.6426×10^{-5} , an improvement of an order of magnitude. Furthermore, DeFINO yields a lower reduced Gauss–Newton error (Equation 5) of 0.7697, compared to 0.8554 for the FNO, highlighting its enhanced ability to approximate critical second-order information. As illustrated in Figure 3, the gradient predicted by DeFINO more closely resembles the structural features of the ground-truth gradient, capturing both upper and lower spherical regions, whereas the FNO prediction reflects only the lower part. Collectively, these results indicate that incorporating derivative-informed training guided by the Fisher Information Matrix substantially improves gradient fidelity without compromising forward prediction accuracy, making DeFINO particularly well-suited for inversion and other gradient-sensitive downstream applications.

DISCUSSION AND CONCLUSION

In this work, we introduced DeFINO, a reduced-order derivative-informed Neural Operator framework designed to enhance surrogate modeling for time-dependent multiphase flow problems. By incorporating derivative-based training via projections onto dominant eigen-directions of the FIM, DeFINO achieves improved accuracy in capturing gradients and Gauss–Newton Hessians, as demonstrated by our results in Table 1. These improvements are particularly notable in scenarios with limited training data, reflecting realistic constraints often encountered in practice. Additionally, our approach reduces the computational cost of derivative information from $O(d^4)$ to $O(r \times d^2)$ by avoiding the explicit formation of full Jacobians. The enhanced derivative estimation and more accurate Hessian approximations provided by DeFINO make it a promising approach for Bayesian inversion, optimization under uncertainty, and robust forecasting of subsurface fluid-flow dynamics.

While these initial results are encouraging, further developments remain necessary for broader applicability and robustness. Future work will explore using the trained DeFINO model within a Bayesian inversion framework, enabling efficient estimation of reservoir permeability fields directly from observational data. Such inverse modeling can enhance uncertainty quantification, helping stakeholders better manage risks associated with CO_2 plume migration. Finally, assessing the generalization capability of DeFINO across diverse subsurface scenarios and extending the framework toward large-scale, continuous 3D monitoring workflows represent promising avenues for future research.

ACKNOWLEDGEMENT

During the preparation of the work, the authors used ChatGPT to refine sentence structures of the manuscript. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

- Benner, P., S. Gugercin, and K. Willcox, 2015, A survey of projection-based model reduction methods for parametric dynamical systems: SIAM review, **57**, 483–531.
- Cai, S., Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, 2021, Physics-informed neural networks (pinns) for fluid mechanics: A review: Acta Mechanica Sinica, **37**, 1727–1738.
- Chandra, A., M. Koch, S. Pawar, A. Panda, K. Azizzadenesheli, J. Snippe, F. O. Alpak, F. Hariri, C. Etienam, P. Devarakota, et al., 2025, Fourier neural operator based surrogates for co_2 storage in realistic geologies: arXiv preprint arXiv:2503.11031.
- Constantine, P. G., 2015, Active subspaces: Emerging ideas for dimension reduction in parameter studies: SIAM.
- Constantine, P. G., E. Dow, and Q. Wang, 2014, Active subspace methods in theory and practice: applications to kriging surfaces: SIAM Journal on Scientific Computing, **36**, A1500–A1524.
- Cui, T., and O. Zahm, 2021, Data-free likelihood-informed dimension reduction of bayesian inverse problems: Inverse Problems, **37**, 045009.
- Gahlot, A. P., R. Orozco, Z. Yin, and F. J. Herrmann, 2024, An uncertainty-aware digital shadow for underground multimodal co_2 storage monitoring: arXiv preprint arXiv:2410.01218.
- Gao, Z., L. Yan, and T. Zhou, 2024, Adaptive operator learning for infinite-dimensional bayesian inverse problems: SIAM/ASA Journal on Uncertainty Quantification, **12**, 1389–1423.
- Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021, Physics-informed machine learning: Nature Reviews Physics, **3**, 422–440.
- Kovachki, N., Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, 2023, Neural operator: Learning maps between function spaces with applications to pdes: Journal of Machine Learning Research, **24**, 1–97.
- Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, 2020a, Fourier neural operator for parametric partial differential equations: arXiv preprint arXiv:2010.08895.
- , 2020b, Neural operator: Graph kernel network for partial differential equations: arXiv preprint arXiv:2003.03485.
- Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, A. Stuart, K. Bhattacharya, and A. Anandkumar, 2020c, Multipole graph neural operator for parametric partial differential equations: Advances in Neural Information Processing Systems, **33**, 6755–6766.
- Li, Z., H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, 2024, Physics-informed neural operator for learning partial differential equations: ACM/JMS Journal of Data Science, **1**, 1–27.
- Lu, L., P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, 2021, Learning nonlinear operators via deeponet based on the universal approximation theorem of operators: Nature machine intelligence, **3**, 218–229.
- Møyner, O., 2024, Jutuldarcy.jl - a fully differentiable high-performance reservoir simulator based on automatic differentiation: **2024**, 1–9.
- O’Leary-Roseberry, T., P. Chen, U. Villa, and O. Ghattas, 2024, Derivative-informed neural operator: an efficient framework for high-dimensional parametric derivative learning: Journal of Computational Physics, **496**, 112555.
- O’Leary-Roseberry, T., U. Villa, P. Chen, and O. Ghattas, 2022, Derivative-informed projected neural networks for high-dimensional parametric maps governed by pdes: Computer Methods in Applied Mechanics and Engineering, **388**, 114199.
- Park, J., N. Yang, and N. Chandramoorthy, 2024, When are dynamical systems learned from time series data statistically accurate?: arXiv preprint arXiv:2411.06311.
- Qiu, Y., N. Bridges, and P. Chen, 2024, Derivative-enhanced deep operator network: arXiv preprint arXiv:2402.19242.
- Quarteroni, A., A. Manzoni, and F. Negri, 2015, Reduced basis methods for partial differential equations: an introduction: Springer, **92**.
- Rosofsky, S. G., H. Al Majed, and E. Huerta, 2023, Applications of physics informed neural operators: Machine Learning: Science and Technology, **4**, 025022.
- Sheriffdeen, S., J. C. Ragusa, J. E. Morel, M. L. Adams, and T. Bui-Thanh, 2019, Accelerating pde-constrained inverse solutions with deep learning and reduced order models: arXiv preprint arXiv:1912.08864.
- Wen, G., Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson, 2022, U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow: Advances in Water Resources, **163**, 104180.
- Wu, K., T. O’Leary-Roseberry, P. Chen, and O. Ghattas, 2023, Large-scale bayesian optimal experimental design with derivative-informed projected neural network: Journal of Scientific Computing, **95**, 30.
- Wu, T., W. Neiswanger, H. Zheng, S. Ermon, and J. Leskovec, 2024, Uncertainty quantification for forward and inverse problems of pdes via latent global evolution: Proceedings of the AAAI Conference on Artificial Intelligence, 320–328.
- Yin, Z., G. Bruer, and M. Louboutin, 2024, slim-group/jutuldarcyrules.jl: v0.2.8.
- Yin, Z., R. Orozco, M. Louboutin, and F. J. Herrmann, 2023, Solving multiphysics-based inverse problems with learned surrogates and constraints: Advanced Modeling and Simulation in Engineering Sciences, **10**, 14.
- Yin, Z., A. Siahkoohi, M. Louboutin, and F. J. Herrmann, 2022, Learned coupled inversion for carbon sequestration monitoring and forecasting with fourier neural operators: Second International Meeting for Applied Geoscience & Energy, Society of Exploration Geophysicists and American Association of Petroleum . . . , 467–472.
- Zahm, O., T. Cui, K. Law, A. Spantini, and Y. Marzouk, 2022, Certified dimension reduction in nonlinear bayesian inverse problems: Mathematics of Computation, **91**, 1789–1835.
- Zhang, K., Y. Zuo, H. Zhao, X. Ma, J. Gu, J. Wang, Y. Yang, C. Yao, and J. Yao, 2022, Fourier neural operator for solving subsurface oil/water two-phase flow partial differential equation: Spe Journal, **27**, 1815–1830.